Merit Functions and Descent Algorithms for a Class of Variational Inequality Problems

Michael Patriksson*

June 15, 2011

Abstract. We consider a variational inequality problem, where the cost mapping is the sum of a single-valued mapping and the subdifferential mapping of a convex function. For this problem we introduce a new class of equivalent optimization formulations; based on them, we also provide the first convergence analysis of descent algorithms for the problem. The optimization formulations constitute generalizations of those presented by Auchmuty [Auc89], and the descent algorithms are likewise generalizations of those of Fukushima [Fuk92], Larsson and Patriksson [LaP94] and several others, for variational inequality problems with single-valued cost mappings.

Key Words. Variational Inequality Problems, Merit Functions, Fenchel's Inequality, Cost Approximation, Descent Algorithms

Mathematics Subject Classification 1991.
Primary: 49J40, 49M05; Secondary: 90C30, 90C33

1 Introduction

Let X be a nonempty, closed and convex subset of $\Re^n$, $F : X \mapsto \Re^n$ a continuous mapping on X, and $u : \Re^n \mapsto \Re \cup \{+\infty\}$ a lower semicontinuous (l.s.c.), proper and convex function. The *variational inequality problem* is to find a pair $(x^*, \xi_u^*) \in X \times \partial u(x^*)$ such that

$$[F(x^*) + \xi_u^*]^{\mathrm{T}} (x - x^*) \geq 0, \qquad \forall x \in X. \tag{1.1}$$

The more general formulation of (1.1) where ∂u is replaced by a general point–to–set mapping was introduced by Fang and Peterson [FaP82] as the *generalized variational inequality*; the problem (1.1) is also known under the name *nonlinear variational inequality* ([Noo91]); through the use of the *normal cone* operator,

$$N_X(x) = \begin{cases} \{z \in \Re^n \mid z^{\mathrm{T}}(y - x) \leq 0, \qquad \forall y \in X\}, & x \in X, \\ \emptyset, & x \notin X, \end{cases}$$

(1.1) can be equivalently written as

$$-F(x^*) \in \partial u(x^*) + N_X(x^*), \tag{1.2}$$

*Department of Mathematics, Linköping University, S-581 83 Linköping, Sweden.

which is frequently referred to as a *generalized equation* ([Rob79]), or a *set inclusion*. The problem (1.1) has a large variety of applications in the mathematical, engineering and social sciences; see, e.g., [EkT76, KiS80, HaP90].

In this paper, we introduce a class of reformulations of (1.1) as equivalent optimization problems, and descent methods based on them. For the special case of variational inequality problems where $u \equiv 0$, there is a plentiful of descent algorithms (see, e.g., [Fuk92, ZhM93, Pat93c, LaP94, Pat94]); the descent algorithms presented in this paper are, however, the first for solving variational inequality problems with point–to–set cost mappings.

We will throughout the paper assume that (1.1) admits at least one solution, and denote the solution set of (1.1) by Ω. The assumptions on X, F, and u made above are assumed to hold throughout the paper.

Sufficient conditions for the non-emptiness of the solution set, Ω, are, for example, that F is monotone on X and that either $\texttt{dom}\, u \cap X$ is bounded [$\texttt{int}\,(\texttt{dom}\, u) \cap X$ assumed nonempty; see Assumption 2.1 below] or that a *coercivity* condition holds, that is, that there is a $z \in \texttt{dom}\, u \cap X$ such that

$$\lim_{\substack{x \in X \\ \|x\| \to \infty}} \left\{ \left[F(x)^{\mathrm{T}} (x - z) + u(x) \right] / \|x\| \right\} = +\infty \tag{1.3}$$

([EkT76, Thm. 3.1]). Moreover, the set Ω is a singleton if it is nonempty and if either F is *strictly monotone* on X, that is, if

$$[F(x) - F(y)]^{\mathrm{T}} (x - y) > 0, \qquad x, y \in X \text{ and } x \neq y,$$

or if u is strictly convex on X. Note that *strong monotonicity* of F (with modulus $m_F > 0$), that is,

$$[F(x) - F(y)]^{\mathrm{T}} (x - y) \geq m_F \|x - y\|^2, \qquad x, y \in X, \tag{1.4}$$

and *strong convexity* of u (with modulus $m_u > 0$), that is, for all $x, y \in X$ and $\lambda \in [0, 1]$,

$$u(\lambda x + (1 - \lambda) y) \leq \lambda u(x) + (1 - \lambda) u(y) - \frac{m_u}{2} \lambda (1 - \lambda) \|x - y\|^2,$$

implies strict monotonicity and strict convexity, respectively, as well as the coercivity condition (1.3), and hence the existence of a unique solution to (1.1).

The rest of the paper is organized as follows. In Section 2, we present a class of optimization formulations of (1.1), which extends that of Auchmuty [Auc89] and those of Larsson and Patriksson [Pat93c, LaP94], which are all concerned with the case $u \equiv 0$. Some interesting properties of the optimization problems are derived. In particular, their derivation reveals a strong relationship between the merit functions and Fenchel's inequality. In Section 3, we present a class of descent algorithms based on these optimization formulations, and establish their global convergence.

2 A class of merit functions for variational inequality problems

It is convenient for the subsequent analysis to reformulate (1.1) into a variational inequality problem containing the convex function u instead of its subdifferential mapping.

This is possible under the following regularity assumption, which is assumed to hold throughout the paper.

Assumption 2.1 (A regularity assumption). $\mathtt{int}\ (\mathtt{dom}\ u) \cap X \neq \emptyset$.

Remark 2.2 The assumption is introduced to ensure that $\partial(u + \delta_X)(x) = \partial u(x) + \partial \delta_X(x)$, $x \in X$, where δ_X is the indicator function of X, and may be replaced by, for example, $\mathtt{rint}\ (\mathtt{dom}\ u) \cap \mathtt{rint}\ X \neq \emptyset$, where $\mathtt{rint}$ denotes relative interior; it can be further weakened whenever u is a polyhedral function or X is polyhedral. See Rockafellar [Roc70, Roc81] for further details on various regularity conditions. □

Proposition 2.3 (Equivalent variational inequality formulation). *Under Assumption 2.1, the problem* (1.1) *is equivalent to the problem of finding an* $x^* \in X$ *such that*

$$F\left(x^*\right)^{\mathrm{T}}\left(x - x^*\right) + u\left(x\right) - u\left(x^*\right) \geq 0, \qquad \forall x \in X. \tag{2.1}$$

Proof. Consider the convex problem

$$\min_{x \in X}\ h(x) \stackrel{\text{def}}{=} F\left(x^*\right)^{\mathrm{T}} x + u(x),$$

where $x^* \in X$. It is clear that (2.1) is equivalent to x^* being a globally optimal solution to this problem. By virtue of Assumption 2.1, we may characterize x^* by the inclusion

$$0 \in \partial h(x^*) + N_X(x^*) \tag{2.2}$$

([Roc70, Thm. 27.4]). Further, Assumption 2.1 implies that

$$\partial h(x) = F(x^*) + \partial u(x), \qquad x \in X; \tag{2.3}$$

combining (2.2) and (2.3) yields that x^* is characterized by the inclusion (1.2), that is, by the variational inequality (1.1). This completes the proof. □

If F is the gradient mapping of some function $f : X \mapsto \Re$, then (2.1) defines the optimality conditions of the constrained minimization problem

$$\min_{x \in X}\ \{f(x) + u(x)\} \tag{2.4}$$

(e.g., [Céa78, Thm. 2.3]), and (2.1) may hence be solved indirectly through (2.4). If neither this is the case, nor is F a mapping of the form $[\nabla_v \Pi(v,w)^{\mathrm{T}}, -\nabla_w \Pi(v,w)^{\mathrm{T}}]^{\mathrm{T}}$ for some saddle function $\Pi : V \times W \mapsto \Re$, then there is no optimization formulation of (2.1) directly available. The purpose of this paper is to construct equivalent optimization formulations of (2.1), and to devise convergent descent algorithms based on them.

The merit functions discussed in this paper are all of the type given by the following definition.

Definition 2.4 (Gap function). *A function* $\psi : X \mapsto \Re \cup \{-\infty, +\infty\}$ *is a* gap function *for* (2.1) *if*

1. ψ *is restricted in sign on* X, *and*

2. $\psi(x) = 0 \Longleftrightarrow x \in \Omega$.

We introduce the function $L : X \times X \mapsto \Re$, with

$$L(x, y) = u(x) - u(y) + \varphi(x) - \varphi(y) + [F(x) - \nabla\varphi(x)]^{\mathrm{T}} (x - y), \qquad x, y \in X,$$

where $\varphi : \Re^n \mapsto \Re \cup \{+\infty\}$ is convex and l.s.c., and $\varphi \in C^1$ on X. Further, we define the function $\psi : X \mapsto \Re \cup \{+\infty\}$ through

$$\psi(x) = \sup_{y \in X} L(x, y), \tag{2.5}$$

and the optimization problem

$$\inf_{x \in X} \psi(x). \tag{2.6}$$

The convex problem defining $\psi(x)$ may be interpreted as follows. A member $y(x)$ of the set, $Y(x)$, of optimal solutions to (2.5) is characterized by the variational inequality

$$[\nabla\varphi(y(x)) + F(x) - \nabla\varphi(x)]^{\mathrm{T}} (y - y(x)) + u(y) - u(y(x)) \geq 0, \qquad \forall y \in X, \tag{2.7}$$

or, equivalently, the inclusion

$$-(F(x) - \nabla\varphi(x)) \in \nabla\varphi(y(x)) + \partial u(y(x)) + N_X(y(x)), \tag{2.8}$$

that is, the problem (2.5) is obtained, at the point $x \in X$, by replacing the mapping F with the monotone gradient mapping $\nabla\varphi$, and adding to this cost mapping the fixed error term $F(x) - \nabla\varphi(x)$. This process is in [Pat93c] referred to as a *cost approximation*, and is a direct extension of that defining the optimization formulations in [Auc89, LaP94] for the special case $u \equiv 0$.

Note that the function L is always concave in y, but not always convex in x. It is therefore not a true saddle function in general. (See Proposition 2.12 for a sufficient condition for the convex–concavity of L and the resulting convexity of ψ.) Note also that L is invariant under the addition of an affine function to φ.

Theorem 2.5 (ψ is a merit function for (2.1)). *ψ is a gap function.*

Proof. In order to prove the theorem, we shall make use of the Fenchel [Fen49] inequality. This result states that

$$h(x) + h^\circ(z) - x^{\mathrm{T}} z \geq 0, \qquad \forall x, z \in \Re^n, \tag{2.9}$$

and that equality holds in (2.9) if and only if $z \in \partial h(x)$, where $h : \Re^n \mapsto \Re \cup \{+\infty\}$ is any proper, convex function, and the conjugate function, h°, of h is

$$h^\circ(w) = \sup_y \{y^{\mathrm{T}} w - h(y)\}. \tag{2.10}$$

To show that Fenchel's inequality yields the desired property of ψ, we identify h with the function $u + \delta_X + \varphi$, and let $z = \nabla\varphi(x) - F(x)$ in (2.9). (The function h is proper from Assumption 2.1, and convex, since it is the sum of convex functions.) From (2.10),

$$\begin{aligned}
\psi(x) &= u(x) + \delta_X(x) + \varphi(x) + \sup_y \left\{ [\nabla\varphi(x) - F(x)]^{\mathrm{T}} y - u(y) - \delta_X(y) - \varphi(y) \right\} \\
&\quad - [\nabla\varphi(x) - F(x)]^{\mathrm{T}} x \\
&= h(x) + h^{\circ}(\nabla\varphi(x) - F(x)) - [\nabla\varphi(x) - F(x)]^{\mathrm{T}} x \\
&= h(x) + h^{\circ}(z) - x^{\mathrm{T}} z, \qquad (2.11)
\end{aligned}$$

which shows that *the value $\psi(x)$ may be interpreted as the gap (or, slack) in Fenchel's inequality (2.9)*, for the special choice of convex function h. From the relationship (2.11), the inequality (2.9) then yields that $\psi(x) \geq 0$ for all $x \in X$, and that

$$\begin{aligned}
\psi(x) = 0 &\iff \nabla\varphi(x) - F(x) \in \partial u(x) + N_X(x) + \nabla\varphi(x) \qquad (2.12) \\
&\iff -F(x) \in \partial u(x) + N_X(x) \\
&\iff x \in \Omega;
\end{aligned}$$

hence, ψ is a gap function for (2.1). □

Remark 2.6 The idea of using Fenchel's inequality to derive merit functions for variational inequality problems originates in [Auc89] for the case $u \equiv 0$. Fenchel's inequality was earlier used in [Mur70] to derive the *primal gap function* (which is obtained from ψ when $u \equiv 0$ and $\varphi \equiv 0$) of a convex program arising in the analysis of traffic networks. Utilizing the properties of convex functions in the context of the above result, it may be fruitful to investigate similar optimization formulations of non-optimization problems other than the class of variational inequality problems studied here. □

From the interpretation of the merit function ψ as *the gap in Fenchel's inequality*, it immediately follows that ψ is l.s.c. on X, since it is the sum of l.s.c. functions. If u is continuous on X, then ψ is continuous on X whenever it is finite on X. Since continuity of ψ is required to obtain closedness of the algorithmic map describing the line search (which is part of the iterative algorithm developed in Section 3), we will henceforth assume that *u is continuous on X*. (Finiteness of ψ is ensured if X is bounded, or if u or φ is strongly convex on X.)

The next result provides a characterization of a solution to (2.1) in terms of a fixed point property of the mapping $x \mapsto Y(x)$ describing the subproblem (2.5).

Proposition 2.7 (A fixed point characterization of Ω)**.** *For any $x \in X$, let $Y(x)$ denote the (possibly empty) set of solutions to (2.5). Then,*

$$x \in \Omega \quad \iff \quad x \in Y(x).$$

Proof. The result follows by the equivalence of (2.12) and (2.8), $y(x)$ replaced by x. $\square$

Further continuity properties of the function ψ, and the mappings $x \mapsto Y(x)$ and $x \mapsto D(x) = Y(x) - x$ are collected in the following result.

Proposition 2.8 (Continuity properties of ψ, Y, and D)**.**

(a) *For any $x \in X$, the set $Y(x)$ is closed and convex, the mappings $x \mapsto Y(x)$ and $x \mapsto D(x)$ are closed, and ψ is l.s.c.*

(b) *If X is bounded or if u or φ is strongly convex on X, then for any $x \in X$, $Y(x)$ is nonempty and bounded, and ψ is continuous on X.*

(c) *If X is bounded and u or φ is strictly convex on X, or if u or φ is strongly convex on X, then, for any $x \in X$, $Y(x)$ is a singleton.*

Proof.

(a) To prove that $x \mapsto Y(x)$ is closed on X, let $\{x^t\}, \{y^t\} \subset X$ be sequences satisfying $\{x^t\} \to x^\infty$, $\{y^t\} \to y^\infty$, and $y^t \in Y(x^t)$ for all t. Letting $t \to \infty$ in the relation

$$-(F(x^t) - \nabla\varphi(x^t)) \in \nabla\varphi(y^t) + \partial u(y^t) + N_X(y^t)$$

[which characterizes the set $Y(x^t)$, cf. (2.8)], by the continuity of F and $\nabla\varphi$ on X, and the closedness of ∂u and N_X ([Roc70, Thm. 24.4]), it follows that

$$-(F(x^\infty) - \nabla\varphi(x^\infty)) \in \nabla\varphi(y^\infty) + \partial u(y^\infty) + N_X(y^\infty),$$

that is, $y^\infty \in Y(x^\infty)$. Hence, the mapping $x \mapsto Y(x)$ is closed on X.

To establish that also $x \mapsto D(x)$ is closed on X, it is sufficient to note that the graph of D is obtained from the graph of Y by an affine transformation. The closedness of the mapping $x \mapsto D(x)$ hence follows from that of $x \mapsto Y(x)$.

The convexity of the set $Y(x)$, $x \in X$, follows from the convexity of the subproblem (2.5) (e.g., [Min62]).

The l.s.c. property of ψ follows from [Hog73, Thm. 6]; see also the above argument.

(b) The first two results are well-known properties of convex problems. The continuity of ψ follows from the continuity and boundedness of L.

(c) The result is a consequence of well-known results for convex problems. $\square$

The next result, and its corollary, extends Theorem 4.5 of [LaP94] to (2.1). A mapping $F : X \mapsto \Re^n$ is *Lipschitz continuous* on X (with modulus $M_F > 0$) if

$$\|F(x) - F(y)\| \leq M_F \|x - y\|, \qquad x, y \in X.$$

Proposition 2.9 (An a posteriori error bound). *Assume that F is strongly monotone on X, and let $\nabla\varphi$ be Lipschitz continuous on X. Let x^* be the unique solution to (2.1). If φ is so chosen that $M_{\nabla\varphi} < m_F$, then*

$$\|x - x^*\|^2 \leq \frac{\psi(x)}{m_F - M_{\nabla\varphi}}, \qquad x \in X. \tag{2.13}$$

Proof. Combining (2.1) with (1.4), with $y = x^*$, yields that

$$F(x)^{\mathrm{T}} (x - x^*) + u(x) - u(x^*) \geq m_F \|x - x^*\|^2, \qquad x \in X. \tag{2.14}$$

For any $y(x) \in Y(x)$,

$$\begin{aligned}
\psi(x) &= u(x) - u(y(x)) + \varphi(x) - \varphi(y(x)) + [F(x) - \nabla\varphi(x)]^{\mathrm{T}} (x - y(x)) \\
&\geq u(x) - u(x^*) + \varphi(x) - \varphi(x^*) + [F(x) - \nabla\varphi(x)]^{\mathrm{T}} (x - x^*) \\
&\geq \varphi(x) - \varphi(x^*) - \nabla\varphi(x)^{\mathrm{T}}(x - x^*) + m_F \|x - x^*\|^2 \\
&\geq (m_F - M_{\nabla\varphi}) \|x - x^*\|^2,
\end{aligned}$$

where the first inequality follows from the fact that $y(x) \in Y(x)$, the second inequality from (2.14), and the third inequality from the convexity of φ and the Lipschitz continuity of $\nabla\varphi$. The proof is complete. □

Corollary 2.10 (Weak coercivity of ψ). *Under the assumptions of the proposition, the function ψ has bounded level sets and is* weakly coercive *on X, that is,*

$$\lim_{\substack{x \in X \\ \|x\| \to \infty}} \psi(x) = +\infty.$$

Proof. The result follows directly from (2.13). □

Remark 2.11 Although the function ψ is non-convex in general, the corollary establishes properties of ψ shared with convex functions. (Weak coercivity and boundedness of level sets are well known properties of *strongly* convex functions.)

The above result enables the methods of this paper to be applied to complementarity type problems, in which case the feasible set X is unbounded (e.g., [Kar71]). □

We next show a case where ψ can be constructed convex (and, simultaneously, the saddle function L convex–concave). This is possible also for non-affine variational inequality problems; we thus give an affirmative answer to the question put in [LaP94, Sec. 4.1].

Proposition 2.12 (Convexity of ψ). *Assume that F is affine, that is, $F(x) = Ax - b$. Let $\varphi(x) = \frac{1}{2}x^{\mathrm{T}}Qx$, where Q is symmetric and positive semidefinite. If the function $x \mapsto u(x) + \frac{1}{2}x^{\mathrm{T}}(A + A^{\mathrm{T}} - Q)x$ is convex on X, then ψ is convex on X.*

Proof. If, for all fixed $y \in X$, the function $x \mapsto L(x, y)$ is convex on X, then the function ψ, which is defined as the pointwise supremum of L, is also convex on X. But the function $x \mapsto L(x, y)$ is the sum of u and a quadratic function, whose Hessian matrix is $A + A^{\mathrm{T}} - Q$. The result then follows immediately. $\square$

The decomposition of the cost mapping into ∂u and F is never unique; adding an arbitrary gradient mapping ∇h of a convex function h to ∂u and subtracting it from F leaves the problem unaltered. Since, in the framework described in this paper, only the mapping F is approximated, this freedom in representing the problem (2.1) leads to the natural question of whether there is something to gain from moving a monotone and additive gradient mapping from F to ∂u, in terms of the convexity properties of ψ, or of the descent properties of the search directions given by the corresponding subproblem solution. This is investigated in the following.

Example 2.13 We first examine the consequences for the convexity properties of the merit function ψ of moving a monotone and additive gradient mapping from F to ∂u. Let F and φ be as in Proposition 2.12. Further, assume that $A = A_1 + A_2$, where A_1 is symmetric, and both A_1 and A_2 are positive semidefinite. Consider two equivalent versions of (2.1), where

(1) u and F are as given, and

(2) $u(x) := u(x) + \frac{1}{2}x^{\mathrm{T}}A_1x$ and $F(x) := F(x) - A_1x$.

We then note that in the first case, the Hessian of the quadratic part of the function $x \mapsto L(x, y)$ is (from Proposition 2.12) $H_1 = A + A^{\mathrm{T}} - Q = 2A_1 + A_2 + A_2^{\mathrm{T}} - Q$, while, in the second case, the Hessian equals $H_2 = A_1 + (A - A_1) + (A^{\mathrm{T}} - A_1) - Q = A_1 + A_2 + A_2^{\mathrm{T}} - Q$. Clearly, the least eigenvalue of H_1 is at least as large as that of H_2, and hence yields a larger modulus of convexity; in terms of the convexity properties of ψ, therefore, there is no gain in moving a monotone and additive gradient mapping from F to ∂u. $\square$

We next provide a calculus rule for the directional derivative of ψ.

Proposition 2.14 (The directional derivative of ψ). *Assume that u is finite on $\Re^n$, and that $F \in C^1$ on X. Let $\varphi \in C^2$ on X. Further, assume that either X is bounded and u or φ is strictly convex on X, or that u or φ is strongly convex on X. Then, for any $x \in X$ and $d \in \Re^n$,*

$$\psi'(x; d) = u'(x; d) + [F(x) + [\nabla F(x)^{\mathrm{T}} - \nabla^2\varphi(x)](x - y(x))]^{\mathrm{T}}d, \tag{2.15}$$

where $y(x)$ is the unique solution to (2.5).

Proof. Under the assumptions of the proposition, it is clear that L is Lipschitz continuous on $X \times X$ since it is the sum of Lipschitz continuous functions (e.g., [Roc81, Proposition 4A]). The assumptions also guarantee that ψ is finite on X (cf. Proposition 2.8.c), and therefore also Lipschitz continuous ([Cla83, p. 92]). In order to invoke

Theorem 2.8.2 of [Cla83] (see also [Roc81, Thm. 5L]), we further note that the graph of ∂L is closed in $\Re^{3n}$, by the Lipschitz continuity of L ([Roc81, Proposition 4R]), and that the directional derivative $L'(x, y; d)$ of L exists for every x, y in X and d in $\Re^n$ (since L is the sum of convex, concave or differentiable functions), so that L is subdifferentially regular ([Roc81, Thm. 4C]). We are then in the position to invoke Theorem 2.8.2 of [Cla83], which (in this special case) states that

$$\partial\psi(x) = \partial_x L(x, y(x)), \qquad x \in X, \tag{2.16}$$

where $\partial_x L(x, y)$ denotes the Clarke generalized gradient of $L(\cdot, y)$ at x. Using this identity, we obtain, for any x in X and d in $\Re^n$, that

$$\begin{aligned} \psi'(x; d) &= \max\{\, z^{\mathrm{T}} d \mid z \in \partial\psi(x) \,\} \\ &= \max\{\, z^{\mathrm{T}} d \mid z \in \partial u(x) + F(x) + [\nabla F(x)^{\mathrm{T}} - \nabla^2\varphi(x)](x - y(x)) \,\}, \end{aligned}$$

where the first equality follows from Theorem 4C of [Roc81], and the second from (2.16) and Theorem 5G of [Roc81]. The formula (2.15) then follows by appealing to Theorem 4C of [Roc81] for the function u. □

If $u \equiv 0$, then the formula (2.15) reduces to that for the merit function given in [LaP94]; the method of proof here is, however, more complicated.

The next proposition characterizes solutions to (2.1) as stationary points to the problem (2.6).

Proposition 2.15 (Stationary point characterization of Ω)**.** *Assume that u is finite on $\Re^n$, and that $F \in C^1$ on X. Let $\varphi \in C^2$ on X. Under either one of the following additional assumptions, x^* solves (2.1) if and only if x^* is a* stationary point *of the problem (2.6), that is, if and only if*

$$\psi'(x^*; x - x^*) \geq 0, \qquad \forall x \in X. \tag{2.17}$$

(1) *X is bounded, u is strictly convex on X, $\nabla F(x^*)$ is positive semidefinite, and φ is quadratic on X.*

(2) *X is bounded, u or φ is strictly convex on X, and $\nabla F(x^*) - \nabla^2\varphi(x^*)$ is positive semidefinite.*

(3) *F is strongly monotone on X or u is strongly convex on X, φ is strongly convex on X, $\nabla\varphi$ is Lipschitz continuous on X, and*

$$M_{\nabla\varphi} - m_\varphi < m_F + \frac{1}{2} m_u. \tag{2.18}$$

Remark 2.16 We make use of the convention that if a mapping T is monotone only, then it satisfies the strong monotonicity defining inequality (1.4) with $m_T = 0$. In the right–hand–side of (2.18), then, (at most) one of the constants m_F and m_u may be zero. Note that for any given strongly monotone mapping F and strongly convex function u the strict inequality (2.18) can be satisfied through a proper scaling of any strongly convex function φ with a Lipschitz continuous gradient; likewise for the (stronger) condition that $M_{\nabla\varphi} < m_F$, which appears in Proposition 2.9. □

Proof of the proposition. That a solution x^* to (2.1) must satisfy (2.17) is obvious.

We next turn to prove the reverse, for the three sets of assumptions, respectively. We first note that the directional derivative $\psi'(x^*; x - x^*)$, for any $x \in X$, equals [cf. (2.15)]

$$\psi'(x^*; x - x^*) = u'(x^*; x - x^*) + [F(x^*) + [\nabla F(x^*)^{\mathrm{T}} - \nabla^2 \varphi(x^*)](x^* - y(x^*))]^{\mathrm{T}}(x - x^*), \tag{2.19}$$

where $y(x^*)$ is the unique solution to (2.5) at x^*.

Using $x = y(x^*)$ in (2.19) we obtain, by assumption, that

$$u'(x^*; y(x^*) - x^*) + [F(x^*) + [\nabla F(x^*)^{\mathrm{T}} - \nabla^2 \varphi(x^*)](x^* - y(x^*))]^{\mathrm{T}}(y(x^*) - x^*) \geq 0. \tag{2.20}$$

Since $y(x^*) \in Y(x^*)$, we have from (2.7) that

$$[\nabla \varphi(y(x^*)) + F(x^*) - \nabla \varphi(x^*)]^{\mathrm{T}}(y - y(x^*)) + u(y) - u(y(x^*)) \geq 0, \qquad \forall y \in X. \tag{2.21}$$

Combining (2.20) and (2.21), with $y = x^*$, we then have that

$$\begin{aligned} 0 \quad \leq \quad & u'(x^*; y(x^*) - x^*) + u(x^*) - u(y(x^*)) \\ & +(x^* - y(x^*))^{\mathrm{T}}[\nabla F(x^*) - \nabla^2 \varphi(x^*)](y(x^*) - x^*) \\ & +[\nabla \varphi(y(x^*)) - \nabla \varphi(x^*)]^{\mathrm{T}}(x^* - y(x^*)). \end{aligned} \tag{2.22}$$

We now turn to the sets of assumptions.

(1) When φ is quadratic on X, (2.22) reduces to

$$0 \leq u'(x^*; y(x^*) - x^*) + u(x^*) - u(y(x^*)) - (y(x^*) - x^*)^{\mathrm{T}} \nabla F(x^*)(y(x^*) - x^*). \tag{2.23}$$

The first three terms of the right–hand–side of (2.23) is non-positive, from the convexity of u, and negative whenever $y(x^*) \neq x^*$ by the strict convexity of u. The last term is non-positive, since $\nabla F(x^*)$ is positive semidefinite. In order for (2.23) to hold, it is thus necessary that $y(x^*) = x^*$. Using this relation in (2.21) then yields that x^* solves (2.1).

(2) The result follows from using a similar line of arguments as that for **(1)**.

(3) Using the monotonicity and convexity properties of F and u, respectively, and the strong monotonicity and Lipschitz continuity properties of $\nabla \varphi$ in (2.22), it follows that

$$\left(\frac{1}{2} m_u + m_F - M_{\nabla \varphi} + m_\varphi \right) \|y(x^*) - x^*\|^2 \leq 0,$$

which, together with (2.18), yields that $y(x^*) = x^*$. As in **(1)**, we may conclude that x^* solves (2.1). $\square$

It is a direct consequence of this result, that if a feasible point x is not a solution to (2.1), then $d = y(x) - x$ defines a feasible direction of descent with respect to ψ. Based on this observation we shall in the next section establish the convergence of descent algorithms that utilize these as search directions.

3 A descent algorithm for variational inequalities

Lemma 3.1 (Descent properties)**.** *Assume that u is finite on $\Re^n$, and that $F \in C^1$ on X. Let $\varphi \in C^2$ on X. Also, let $x \in X \setminus \Omega$, and $y(x)$ be the solution to (2.5).*

(a) *Assume that X is bounded, u is strictly convex on X, and $\nabla F(x)$ is positive semidefinite. Let φ be quadratic on X. Then,*

$$\psi'(x; y(x) - x) < 0 \tag{3.1}$$

holds.

(b) *Assume that X is bounded, u or φ is strictly convex on X, and $\nabla F(x) - \nabla^2\varphi(x)$ is positive semidefinite. Then, (3.1) holds.*

(c) *Assume that F is strongly monotone on X or u is strongly convex on X. Let φ be strongly convex on X and $\nabla\varphi$ Lipschitz continuous on X. Then,*

$$\psi'(x; y(x) - x) \le -\left(\frac{1}{2}m_u + m_F + m_\varphi - M_{\nabla\varphi}\right) \|y(x) - x\|^2 \tag{3.2}$$

holds.

Proof. All the results follow from expressing $\psi'(x; y(x) - x)$ as in (2.19), using its estimate (2.22), and the same techniques as is in the proof of Proposition 2.15. □

Example 3.2 We continue Example 2.13, and investigate the consequences for the descent properties of the search directions of moving a monotone and additive gradient mapping from F to ∂u. Let F and φ be as in that example. Further, assume that u is strongly convex, and that A_1 and A_2 are positive definite. Consider the two equivalent versions of (1.1) given in Example 2.13.

The inequality (2.15) yields that the directional derivative of ψ at x in the direction of $y(x) - x$ in the first alternative satisfies

$$\psi'(x; y(x) - x) \le -\left(\frac{1}{2}m_u + m_F\right) \|y(x) - x\|^2,$$

while

$$\psi'(x; y(x) - x) \le -\left[\frac{1}{2}(m_u + m_{F_1}) + m_{F_2}\right] \|y(x) - x\|^2$$

in the second alternative, where $F_1(x) = A_1 x$ and $F_2(x) = A_2 x - b$. Clearly,

$$\frac{1}{2}m_u + m_F \ge \frac{1}{2}(m_u + m_{F_1}) + m_{F_2},$$

and, disregarding the fact that the search directions in the two alternatives may not have the same length, the first version yields steeper directions; in any event, the condition (2.18) is milder in the first case. In Example 2.13, our conclusion was that there was nothing to gain from moving a monotone and additive gradient mapping from F to ∂u, and the conclusion here regarding the descent properties is the same. This analysis can be summarized by the conclusion that all the strong monotonicity inherent in the problem mapping should be kept in F. The result thus reached can be put to good use for improving the property of both the equivalent optimization problem and the algorithm used for solving it. $\square$

We have shown that the solution to (2.5), obtained when evaluating $\psi(x)$, defines a feasible descent direction with respect to ψ, whenever $x \notin \Omega$. An iterative algorithm, which combines the solution of (2.5) with an exact line search with respect to the merit function ψ, is validated below.

Theorem 3.3 (Global convergence)**.** *Assume that u is finite on $\Re^n$, and that $F \in C^1$ on X. Let $\varphi \in C^2$ on X. Let x^0 be arbitrary in X, and x^{t+1} be determined from x^t through the exact line search rule*

$$\min_{0 \leq \ell \leq 1} \psi\left(x^t + \ell d^t\right),$$

where $d^t = y(x^t) - x^t$. Consider the following sets of assumptions.

(1) *X is bounded, u is strictly convex on X, ∇F is positive semidefinite on X, and φ is quadratic on X.*

(2) *X is bounded, u or φ is strictly convex on X, and $\nabla F - \nabla^2 \varphi$ is positive semidefinite on X.*

(3) *F is strongly monotone on X or u is strongly convex on X, φ is strongly convex on X, and $\nabla\varphi$ is Lipschitz continuous on X. Further, (2.18) holds, and either X is bounded or F is strongly monotone and $M_{\nabla\varphi} < m_F$ holds.*

Under either one of the above sets of assumptions, (2.1) has a unique solution, the sequence $\{x^t\}$ converges to it, and the sequence $\{\psi(x^t)\}$ strictly monotonically decreases to zero.

Proof. We begin by showing that the assumptions of Theorem A of Zangwill [Zan69, Sec. 4.5]) are fulfilled in all the three cases, that is, *i*) that the sequence $\{x^t\}$ stays in a compact subset of X, *ii*) that $\psi(x^{t+1}) < \psi(x^t)$ whenever $x^t \notin \Omega$, and $x^t \in \Omega$ if $\psi(x^{t+1}) \geq \psi(x^t)$, and *iii*) that the algorithmic map is closed.

i) For cases **(1)** and **(2)**, this is obvious since X is compact. For case **(3)**, either it holds from the compactness of X or from the compactness of the level sets (cf. Corollary 2.10).

ii) Follows directly from Proposition 2.15 and Lemma 3.1, and the choice of step length.

iii) The closedness of the search direction finding map follows from Proposition 2.8.a. The line search map is closed since ψ is continuous on X. The composite mapping is hence closed on the compact intersection of X with the level set of ψ at x^0 in all cases (cf. Corollary 4.2.1 of [Zan69]).

Thus, by Zangwill's Theorem A, every accumulation point [from *i*), at least one such point exists] is stationary in (2.6). In all the three cases **(1)**–**(3)**, Ω is a singleton, and by Proposition 2.15 we may conclude that $\{x^t\}$ converges to the unique solution of (2.1). $\square$

Remark 3.4 When $u \equiv 0$, the cases **(2)** and **(3)** correspond to the convergence conditions for the descent methods of [LaP94]. The case where φ is quadratic [case **(1)**] is, in [LaP94], not analyzed separately, but that special case of descent method corresponds to that of Fukushima [Fuk92]. His convergence results require φ to be strictly convex; such a condition is not present in case **(1)** in the presence of u. $\square$

Acknowledgements

The author would like to thank the referees, Associate Professor T. Larsson and Professor R.T. Rockafellar for helpful comments. The research leading to this paper was partially supported by grants from the Swedish Institute (303 GH/MLH), the Royal Swedish Academy of Sciences, and the Swedish Research Council for Engineering Sciences (TFR) (282-93-1195).

References

[Auc89] Auchmuty, G., *Variational principles for variational inequalities,* Numerical Functional Analysis and Optimization, 10 (1989) 863–874.

[Céa78] Céa, J. *Optimization: Theory and Algorithms,* Springer-Verlag, Berlin, 1978.

[Cla83] Clarke, F. H., *Optimization and Nonsmooth Analysis,* John Wiley & Sons, New York, NY, 1983.

[EkT76] Ekeland, I., and Temam, R., *Convex Analysis and Variational Problems,* North-Holland, Amsterdam, 1976.

[FaP82] Fang, S.-C., and Peterson, E.L., *Generalized variational inequalities,* Journal of Optimization Theory and Applications, 38 (1982) 363–383.

[Fen49] Fenchel, W., *On conjugate convex functions,* Canadian Journal of Mathematics, 1 (1949) 73–77.

[Fuk92] Fukushima, M., *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems,* Mathematical Programming, 53 (1992) 99–110.

[HaP90] Harker, P. T., and Pang, J.-S., *Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications,* Mathematical Programming, 48 (1990) 161–220.

[Hog73] Hogan, W. W., *Point-to-set maps in mathematical programming,* SIAM Review, 15 (1973) 591–603.

[Kar71] Karamardian, S., *Generalized complementarity problem,* Journal of Optimization Theory and Applications, 8 (1971) 161–168.

[KiS80] Kinderlehrer, D., and Stampacchia, G., *An Introduction to Variational Inequalities and Their Applications,* Academic Press, New York, NY, 1980.

[LaP94] Larsson, T., and Patriksson, M., *A class of gap functions for variational inequalities,* Mathematical Programming, 64 (1994) 53–79.

[Min62] Minty, G. J., *Monotone (nonlinear) operators in Hilbert space,* Duke Mathematical Journal, 29 (1962) 341–346.

[Mur70] Murchland, J. D., *Road network traffic distribution in equilibrium,* In: Mathematical Models in the Social Sciences, II Oberwolfach-Tagung über Operations Research, Mathematisches Forschungsinstitut, Oberwolfach, 20–25 October 1969, R. Henn, H. P. Künzi, and H. Schubert (eds.), Anton Hain Verlag, Meisenheim am Glan, West Germany, pp. 145–183, 1970. In German, translation by H. A. Paul.

[Noo91] Noor, M. A., *Some classes of variational inequalities,* In: Constantin Carathéodory: An international tribute, Th. M. Rassias (ed.), World Scientific Publ. Co., Singapore, pp. 996–1019, 1991.

[Pat93a] Patriksson, M., *Partial linearization methods in nonlinear programming,* Journal of Optimization Theory and Applications, 78 (1993) 227–246.

[Pat93b] Patriksson, M., *A unified description of iterative algorithms for traffic equilibria,* European Journal of Operational Research, 71 (1993) 154–176.

[Pat93c] Patriksson, M., *A unified framework of descent algorithms for nonlinear programs and variational inequalities,* Ph.D. dissertation, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden, 1993.

[Pat94] Patriksson, M., *On the convergence of descent methods for monotone variational inequalities,* Operations Research Letters, 16 (1994) 265–269.

[Rob79] Robinson, S. M., *Generalized equations and their solutions, part I: basic theory,* Mathematical Programming Study, 10 (1979) 128–141.

[Roc70] Rockafellar, R. T., *Convex Analysis,* Princeton University Press, Princeton, NJ, 1970.

[Roc81] Rockafellar, R. T., *The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions,* Heldermann Verlag, Berlin, 1981.

[Zan69] Zangwill, W. I., *Nonlinear Programming: A Unified Approach,* Prentice-Hall, Englewood Cliffs, NJ, 1969.

[ZhM93] Zhu, D. L., and Marcotte, P., *Modified descent methods for solving the monotone variational inequality problem*, Operations Research Letters, 14 (1993) 111–120.